



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/934,223	08/21/2001	YeYi Wang	M61.12-0360	8169
27366	7590	06/17/2005	EXAMINER	
MICROSOFT CORPORATION C/O WESTMAN CHAMPLIN & KELLY, P.A. SUITE 1400 - INTERNATIONAL CENTRE 900 SECOND AVENUE SOUTH MINNEAPOLIS, MN 55402-3319			HAN, QI	
			ART UNIT	PAPER NUMBER
			2654	
DATE MAILED: 06/17/2005				

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)
	09/934,223	WANG, YEYI
	Examiner	Art Unit
	Qi Han	2654

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) Responsive to communication(s) filed on _____.
- 2a) This action is FINAL. 2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) Claim(s) 1-28 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) Claim(s) _____ is/are allowed.
- 6) Claim(s) 1-28 is/are rejected.
- 7) Claim(s) 20 is/are objected to.
- 8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on _____ is/are: a) accepted or b) objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) All b) Some * c) None of:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. _____.
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date. _____ |
| 3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date <u>7/11/02 & 1/18/05</u> | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

Information Disclosure Statement

1. The references listed in the Information Disclosure Statement submitted on 07/11/2002

and 01/18/2005 have been considered by the examiner (see attached PTO-1449).

Undated references were not considered. DR

Specification and Drawings

2. The disclosure is objected to because of the following:

- a. On page 20, lines 14-16, the content of specification say “The hash table **maps a token ID to an array** of partial parses that need that token ID to be extended”.

However, Fig. 5 shows that the hash table **maps “token” to “to token ID”**, which is in conflict with the specification. Appropriate correction is required.

- b. On page 20, line 20 to page 21, line 16, even though the specification says “Fig. 8 provides an example of a partial parse hash table 800 for the word “meeting” in the input text...” However, the following description and Fig. 8 do not show any meaningful relationship between the word “meeting” and the hash table at all. For example, the word “meeting” itself is a token, so that it is unclear what relationship is between the token for “meeting” and other “tokens A, B, C and D”. Appropriate correction/explanation is required.

Claim Objections

3. Claim 20 is objected to because of the following informalities:

Regarding claim 20, the terms “staring position” in line 12 of the claim appears to be -- starting position--. Appropriate correction is required

Claim Rejections - 35 USC § 101

35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

4. Claims 1 and 3 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

Regarding claim 1, it recites “method..., selecting a token; identifying an integer...; and utilizing the integer to identify at least one token of the logical representation...”, which is treated as mere arrangement of data, wherein “token” and “integer” are interpreted as pure data. Since the claimed limitation lacks structurally and functionally interrelationship to any hardware and/or software functionality embedded in a computer related structure/device (and without any computer related post-process and preprocess), the result of the processing lacks a practical application, thus the claimed limitation is directed to non-statutory subject matter.

Even though the preamble recites ‘the method of parsing text to form a logical representation of the text, the logical representation having token representing non-terminal and words of the text’, wherein “text” and “word” are treated as abstract data, it does not change the nature that the claimed method is merely arranging data as stated above, which is directed to non-statutory subject matter.

Regarding claims 2-3 (depending on claim 1), the rejection is based on the same reason describe for claim 1, because claims 2-3 has the same or similar problem regarding non-statuary

subject matter as claim 1.

5. To expedite a complete examination of the instant application the claims rejection under 35 U.S.C 101 (nonstatutory) above are further rejected as set forth below in anticipation of applicant amending these claims to place them within the four statutory categories of invention.

Claim Rejections - 35 USC § 102

5. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

6. Claims 1-10 are rejected under 35 U.S.C. 102(e) as being anticipated by Call (US 2002/0165707 A1).

As per **claim 1**, Call discloses method and apparatus for storing and processing natural language text data as a sequence of fixed length integers (title), comprising:

“selecting a token” (paragraph 11, ‘parsing the text data into logical subdivisions (sub-string of words or tokens) consisting of the alphanumeric term (word or token); paragraphs 41-45, ‘converting terms into corresponding integer tokens’, wherein necessarily including selecting a token);

“identifying an integer that represents the selected token” (Fig. 1 and paragraphs 27-28, ‘each term (token) identified by parser 115 is compared at 121 with the content of a lookup table

125' that 'takes the form of a binary tree data structure', 'performs a binary search... and returns the integer'); and

"utilizing the integer to identify at least one token of the logical representation that begins with the selected token" (paragraph 91, 'data array which hold a termnumber (integer) identifying one of the target terms in wordlist'; Fig.2 shows the structure including logical representation that begins with the token).

As per **claim 2** (depending on claim 1), Call further discloses "identifying an integer that points to an identifier array, each cell in the identifier array providing a token identifier for a token that begins with the selected token" (Fig. 2 and Table 1, shows that the structure comprising 'termnumber (integer)', each row (corresponding to cell) of 'L R O' and the associated array 'T' (including a token identifier, starting and ending position information)).

As per **claim 3** (depending on claim 2), as stated above, Call discloses that "the token identifiers are integers" (paragraphs 41-45, 'converting terms into corresponding integer tokens'; Fig. 2 and Table 1, indicates other related information (identifiers) are integers).

As per **claim 4** (depending on claim 3), Call further discloses that "each token identifier integer comprises a table identifying portion and an offset portion, the table identifying portion identifying a table that contains an array of definitions for tokens including the token represented by the token identifier integer and the offset portion identifying the location of the definition for the token represented by the token identifier integer" (Figs. 2-3 and paragraphs 76- 99, 'the array 310 is identified by the symbolic name "data" (interpreted as table identifying portion)', 'TableA' and 'TableB' that include termnumber (integer) associating definitions based on the structure (array) of 'T 222' in Fig. 2, and 'the value in the array cell at the index location j (offset

portion)' that associates (identifies) a row of 'LRO' in Fig. 2, which provides location information for the token).

As per **claim 5** (depending on claim 4), Call further discloses that "each cell in the identifier array further provides an indication of a rule in which the token represented by the token identifier integer begins with the selected token" (Figs. 2, Table 1 and paragraphs 44-48, wherein the 'termnumber' associates each element (cell) in structures of 'L', 'R' and 'O', which provides definition of the data structure, including indication of rule for token identifier beginning with, such as using root table, starting position information in 'O') .

As per **claim 6**, it recites a computer-readable medium having a data structure. The rejection is based on the same reason as described for claim 1, because the claim recites the same or similar limitations as claim 1.

As per **claim 7** (depending on claim 6), the rejection is based on the same reason as described for claim 3, because the claim recites the same or similar limitations as claim 3.

As per **claim 8** (depending on claim 7), Call further discloses that "the token identifier points to a token definition for a token" (Fig. 2, shows that the related information (identifiers) points to the corresponding substring (token) of 'T 222', which is a token definition).

As per **claim 9** (depending on claim 8), Call further discloses that "the token definition for a token comprises a sequence of token identifiers that can be parsed to form the token defined by the token definition" (Fig. 2 and Table 1, shows that 'termnumber' is related to the corresponding elements in the 'L' 'R' 'O' and 'T', which are interpreted as a sequence of token identifiers and form the token definition).

As per **claim 10** (depending on claim 9), Call further discloses that “each cell in the array further comprises a pointer to a sequence of tokens in the token definition” (Fig. 2 and Table 1, shows that each termnumber (corresponding to cell) is associated a sequence of characters (each of them is also interpreted as a token)).

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

7. Claims 11, 13-15, 17-18 and 20-26 are rejected under 35 U.S.C. 103(a) as being unpatentable over Call (US 2002/0165707 A1) in view of Brash (US 5,960,384).

As per **claim 20**, Call discloses method and apparatus for storing and processing natural language text data as a sequence of fixed length integers (title), comprising:

“identifying a first structure that spans a first sub-string of words in the text and has a first token as its root, the first sub-string having a starting position and an ending position” (paragraph 11, ‘parsing (identifying) the text data into logical subdivisions (sub-string of words or tokens) consisting of the alphanumeric term (word or token)’; paragraphs 41-45, ‘using a binary data structure’, ‘integer tokens’, ‘pre-allocated, vectored binary tree’, ‘parsing process subdivides that input text into the substrings’; Fig. 2, shows that the structure has a first token ‘M’ in Root table 250 and first substring that spans word ‘Mr’ having a starting position

indicated by an offset ‘0’ and ending position indicated by a length ‘2’ in tables ‘O 224’ and ‘T 222’);

“indexing the first structure by the first token and the start(ing) position and ending position of the first sub-string” (paragraphs 44-45, ‘indexed by a termnumber (as a token)’; Fig. 2 and Table 1, shows the first termnumber and the related starting and ending positions of the first sub-string ‘Mr’);

“identifying a second structure that spans the first sub-string of words and has the first token as its root” (paragraph 54, ‘a separate tree is created for all terms (sub-strings) beginning with the same character (first token as its root)’, wherein each path in the tree corresponds to a term that is associated with a structure represented as a row of ‘L R O’ in Fig. 2, which means the system is capable of implementing the claimed limitation, such as treating ‘Mr’ and ‘Mr.’ as separate substring, as suggested by Call in paragraph 26).

“using the first token and the start position and end position of the first sub-string to locate the first structure” (paragraph 32, ‘translating (using) each integer ... back into a character string form using the term lookup table’, wherein necessarily uses the starting and ending position information).

But, Call does not expressly disclose “removing one of the first structure and second structure from further consideration in the formation of the representation of the text”. However, these features are well known in the art as evidenced by Brash who, in the same field of endeavor, discloses method and device for parsing natural language sentences and other sequential symbolic expressions (title), and teaches that ‘some parsers prune (remove) the parallel parsing paths by using syntactic rules to assess the likelihood’ (column 3, lines 60-61).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Call by specifically providing extra parsing paths and pruning some parsing path based on likelihood, as taught by Brash, for the purpose of improving parse correctness (Brash: column 3, lines 61-62).

As per **claims 21 and 22** (depending on claim 20), the rejection is based on the same reason as described for claim 20, because the rejection for claim 20 covers the same or similar limitation(s) of claims 21-22 (see claim 20).

As per **claim 23** (depending on claim 22), as stated above, Call in view of Brash is capable of implanting the claimed functionality “removing the first structure comprises removing the first structure so that it is no longer indexed by the first token and the starting position and ending position of the first sub-string and indexing the second structure by the first token and the starting position and ending position of the first sub-string” (Brash: column 3, lines 60-61, ‘prune the parallel parsing paths’ that suggests no longer indexing those pruned paths; Call: Fig. 2, Table 1 and paragraphs 44-45, ‘indexed by a termnumber (a token)’ and the associated starting and ending position information, for example, as stated above, using ‘Mr’ and ‘Mr.’ as separate substrings, pruning ‘Mr’ and keeping ‘Mr.’ with the associated starting and ending position information).

As per **claim 24** (depending on claim 20), the rejection is based on the same reason as described for claim 20, because the rejection for claim 20 covers the same or similar limitation(s) of claim 24 (see above), wherein ‘assess the likelihood’ (Brash: column 3, lines 60-61) reads on the claimed “comparing … to determine which structure is better for the representation of the text”.

As per **claim 25**, it recites a computer-readable medium having a data structure. The rejection is based on the same reason as described for claim 20, because the claim recites the same or similar limitations as claim 20.

As per **claim 26** (depending on claim 25), Call does not expressly disclose the address field comprising a starting point and an ending point for a set of words in a text string". However, these features are well known in the art as evidenced by Brash who, further discloses 'phrase structure' using 'verb phrase' and/or 'noun phrase' (a set of words) for the parser (column 3, lines 14-31, and column 16, lines 1-2). Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Call by specifically providing a capability for handle some phrase as a substring for a parser, as taught by Brash, for the purpose of improving efficiency for parsing (Brash: column 3, line 31).

As per **claim 11**, the rejection is based on the same reason as described for claim 20, because the rejection for claim 20 covers the same or similar limitations of claim 20, wherein processing the text 'Sam went to' in block 240 of Fig. 2 in Call's reference reads on the claimed "partial parse" and 'parallel parsing paths by using syntactic rules to assess the likelihood' in Brash's reference reads on the claimed "to extend the parse".

As per **claim 13** (depending on claim 11), Call in view of Brash further discloses "identifying an item comprises identifying a word" (Call: Fig. 2, shows the item including word).

As per **claim 14** (depending on claim 11), Call in view of Brash further discloses "identifying an item comprises identifying a non-terminal" (paragraphs 43-44, 'binary tree' that necessarily include non-terminal).

As per **claims 15, 17-18**, they recite a computer-readable medium having a data structure.

The rejection is based on the same reason as described for claims 11 and 13-14, respectively, because the claims recite the same or similar limitation(s) as claims 11 and 13-14, respectively.

8. Claims 12, 16 and 19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Call in view of Brash as applied to claim 11, and further in view of Messerly et al. (US 5,960,384) hereinafter referenced as Messerly.

As per **claim 12** (depending on claim 11), even though Call discloses “placing a pointer to an array in the table” (Fig. 2 and Table 1), Call in view of Brash does not expressly teach that “the array contains at least two partial parses that can be extended by a same item”. However, the features is well known in the art as evidenced by Messerly who discloses that ‘information retrieval utilizing semantic representation of text’ (title), comprising using two partial logical forms (tokens) for combination of matching tokens (Figs. 17-18 and column 12, lines 6-39). Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Call in view of Brash by specifically providing using two matched partial logical forms for the parser, as taught by Messerly, for the purpose of improving tokenizer for parsing input text (Messerly: column 2, lines 35-36).

As per **claim 16** (depending on claim 15), the rejection is based on the same reason described for claim 12 because the claims recite the same or similar limitation(s) as claim 12.

As per **claim 19** (depending on claim 15), the rejection is based on the same reason described for claim 12 because the claims recite the same or similar limitation(s) as claim 12.

9. Claims 27-28 are rejected under 35 U.S.C. 103(a) as being unpatentable over Call in view of Bennett et al. (US 6,615,172B1) hereinafter referenced as Bennett.

As per **claim 27**, Call discloses method and apparatus for storing and processing natural language text data as a sequence of fixed length integers (title), comprising:

“converting a selected token into a token ID” (Fig. 1 and paragraph 16, ‘each parsed substring (selected token) is then converted ...into a binary value (ID)to form an array of integers’; paragraph 76, ‘the value Data[j] (token ID’);

“using a first portion of the token ID to identify a table containing definitions for tokens [of a same type as the selected token]”, (Figs. 2-3 and paragraphs 76- 99, ‘the array 310 is identified by the symbolic name “data” (first portion)’, ‘TableA’, ‘TableB’, wherein TableA or TableB can be broadly interpreted as table containing definition, since they are based on the structure (array) of ‘T 222’ in Fig. 2, that contains definition for tokens),

“using a second portion of the token ID to locate the definition for the selected token in the identified table” (Figs. 2-3 and paragraphs 76-99, ‘the value in the array cell at the index location j (second portion)’, ‘TableA’, ‘TableB’);

“using the definition for the selected token as part of the method of identifying the parse structure” (Figs 2-3, as stated above).

But, Call does not expressly teach definitions for the tokens being of “a same type as the selected token”. However, the features is well known in the art as evidenced by Bennett who discloses that ‘the tokenizer output lists the offset and category (same type)’ and ‘determines which groups of words form phrases (same type of tokens is the group)’ (column 17, lines 48-64). Therefore, it would have been obvious to one of ordinary skill in the art at the time the

invention was made to modify Call by specifically providing pruning some parsing path, as taught by Bennett, for the purpose of improving accuracy for a wider group of users (Bennett: column 5, lines 64-65).

As per **claim 28**, it recites a computer-readable medium having a data structure. The rejection is based on the same reason as described for claim 27, because the claim recites the same or similar limitations as claim 27.

Conclusion

10. Please address mail to be delivered by the United States Postal Service (USPS) as follows:

Mail Stop _____
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

or faxed to:

(703) 872-9306, (for formal communications intended for entry)

Or:

(703) 872-9306, (for informal or draft communications, and please label
"PROPOSED" or "DRAFT")

If no Mail Stop is indicated below, the line beginning Mail Stop should be omitted from the address.

Effective January 14, 2005, except correspondence for Maintenance Fee payments, Deposit Account Replenishments (see 1.25(c)(4)), and Licensing and Review (see 37 CFR 5.1(c) and 5.2(c)), please address correspondence to be delivered by other delivery services (Federal Express (Fed Ex), UPS, DHL, Laser, Action, Purolater, etc.) as follows:

U.S. Patent and Trademark Office
Customer Window, Mail Stop _____
Randolph Building
Alexandria , VA 22314

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Qi Han whose telephone numbers is (703) 305-5631. The examiner can normally be reached on Monday through Thursday from 9:00 a.m. to 7:00 p.m. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil, can be reached on (703) 305-9645.

571-272-7602

571-272-7604

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Inquiries regarding the status of submissions relating to an application or questions on the Private PAIR system should be directed to the Electronic Business Center (EBC) at 866-217-9197 (toll-free) or 703-305-3028 between the hours of 6 a.m. and midnight Monday through Friday EST, or by e-mail at: ebc@uspto.gov. For general information about the PAIR system, see <http://pair-direct.uspto.gov>.

QH/qh
June 6, 2005



**DAVID D. KNEPPER
PRIMARY EXAMINER**